Statewide Data Program Needs Assessment

Bay Delta Office
Francis Chung, Traq Kadir. Hong Bing, Tara Smith, Nicky Sandhu, Eli Ateljevich

February 16, 2011

Program Description

The Bay Delta Office work on statewide system operations for the State Water Project as well as federal projects; they work on current operations (with the Division of Operations and Maintenance) as well as future scenarios.

Most of what they do is planning studies. For engineering analyses, the Office relies on consultants.

The Office works on reliability of State Water Project deliveries.

Half of the funding is from the State Water Project. Half of the funding is from other programs, which the Office is treated like a cost center.

CalSIM
For surface storage studies, the Office uses CalSIM. To understand climate change, the Office needs stream flow data, watershed information and precipitation.

For delivery studies, the Office needs surface water data, hydrology, flood data, areas, water year types, field data, estimates of evapotranspiration, diversions, return flows, soil characteristics, population and estimates of water use efficiency.

The Office works with NASA and CBAL (?) to estimate actual evapotranspiration.

For the State Water Project, the Office models water flows through the Delta for the 29 State Water Contractors. The bathymetry of the Delta and surface water stage are important.

CalSIM3
CalSIM3 is a planning model for all of the Central Valley. This model needs groundwater level data, well construction information. They also use information from Agricultural Commissioners. The model uses 130 demand units. The Office is trying to

model the source and use for each of these 130 spatial areas. Use is limited to some historical data. The model needs to estimate current and future use. The accuracy of the model depends on the data. The model uses stream flows on a daily basis. The old model used to run on a monthly time-step. Now they are converting the model to run on a daily time-step. Francis says there is pressure to convert the model to an hourly time-step.

CalSIM3 uses data from other models, flood data and field data. How frequently the data is collected varies.

CalSIM3 stores data in a DSS format. This was developed by the Army Corps of Engineers. The model produces a lot of data (Gigabytes, if not Terabytes). There is a need to store this information. There is also a need to visualize the information in ArcGIS and CUAHSI.

Francis (or his staff) can email me sources of data for CalSIM3.

Integrated Surface Water and Groundwater
Is this CalSIM3?

This is just a model of water quantity. There are historical and projected quantities, from 1922. It is a simple water balance model of supply and demand. The model used to be a monthly time-step; now it is going to a daily time-step. The Office relies on data sources and the regional offices.

The model is starting to incorporate and model the upper watersheds.

The model uses observed and processed data (aerial data). Data used includes:

❖ Land use data. Historical land use data would be great. If we could digitize the maps in the regional offices that would be helpful.

❖ Regional reports and studies to refine the model.

❖ Precipitation. The Office buys PRISM data (Raster precipitation data) from Oregon State (Go Beavers!).

❖ Evapotranspiration data,

❖ Diversion data is essential, and hard to get.

❖ Groundwater pumping, well data and efficiency estimates. Geology data of the wells and aquifers.

It would be nice to collaborate with PG&E and use their data. But PG&E is afraid of liability and will not give out its data.

Data is stored as part of the model.

70% of the Office's effort for CalSIM3 is gathering and processing the data so it that it can be used in the model.

Francis, Kamyar, Andrew Schwartz and Gary Bardini are looking at how to incorporate climate change into the models.

It would be nice to have a list of the types of data, and the methods of data analysis.

Nicki said that there are issues with data collection. One of which is that the data is not QA/QC'd. Models provide a good evaluation of the data.

Data workflows are an issue. The Office has to create many importers (code) for different data sources. It gets worse, because a single data source is not always consistent, so then exceptions have to be programmed.

Small databases do not work for storing the results of the models.

Model outputs are generated that no one maintains. Then people ask what the model input was three years later. It is hard to know.

There should be a tool to access the data.

Delta Model
The Delta Model, or REALM, is used to forecast and generate planning scenarios. The model uses historical information for calibration. The modelers use CDEC and IEP as data sources in the Department. The model uses LIDAR, consumptive use estimates, island geometry, seepage estimates, surface water and water quality measurements.

Eli Ateljevich
He works on the Delta bathymetry model.

Consumptive use in the Delta islands is a big issue. No one knows what it really is. He is working with NOAA, USGS and DWR to figure it out.

Brad Tom has the Delta island data. He maintains integrated data sets from many agencies.

There should be some data organization of fluxes in the Delta. We just do not know all of the fluxes we need to know to adequately model the Delta.

CDEC has a problem of time shifts for their time series. Day-light saving time/standard time is a big issue.

CDEC site locations are also an issue. The site locations are not very accurate.

There are a lot of hidden treasures of data.

NASA has a web site for their C2 Framework. This is a wiki of different data sets. For large data sets, use NETCDF format, or HDF5. NASA has a formal protocol storing and sending for time series data across the internet.

Recommendations
Recommendations that Curtis and I came up with for the Report we will produce from these surveys.

There should be a one-stop shop for data. Right now, there is not.
The data should be put into an electronic format so that it is easily accessible.
Reduce the duplication of data collected. There is no way to identify who collects what.
DWR needs to find a way to address historical data.
What are the parameters that are measured at a given site?
Need to know that area/region of California some data is pertinent to.
Where is data stored
What is the time period of the data?
What is the accuracy of the data?
What is the quality of the data?
Is metadata available?
What is the form of the data?
Is the data restricted? Or is the data freely available?

Apportioning Costs

Not discussed because there was not time.

Other Issues

Not discussed.